

3G-3

英語文型選択の並列化について

末松 博 村木 一至

日本電気(株) C & C 情報研究所

1. はじめに

文型処理の行々化は、源言語の文の表層から概念依存構造を記述した中間表現へのマッピング、更に中間表現から目的言語の文へのマッピングに非常に重要な役割を担う。解析においては入力文の統語構造解析、用言の意味解析、文法格認定、文法格アキメントに対する意味解析に役立つ。生成においては出力用言選択、深層格アキメントに対する文法格選択、表層語選択に役立つ。

一般的にある用言表層に対し複数の意味が対応し、更に各意味に依存して複数の文型が対応する。又、意味から見ても同様の関係が認められる。従って、解析・生成の両方において文型処理は可能な複数の文型から妥当なものを絞り込む必要がある。

2. 既存の文型処理行々における問題点

既存の文型処理行々としては次の二つの選択法が知られている。

(1) 可能な文型から一つを仮定して選び、順番にそのアキメントの素性をアキメント候補の素性と比較して行き、合致しない段階で別の文型を選び、失敗が検出されなかった文型をその用言の文型と認定する[1]。

(2) 可能な文型を並べて、アキメント候補の素性を全てのアキメント素性と比較し、最多一致した文型を用言の文型と認定する。

(1)においては、選ばれる文型が、仮定される順番に依存し、アキメントとして認定されるべき候補が見落とされる可能性がある。又、アキメントの多い文型から仮定すると、アキメント数の少ない文型の方が頻度が高いために効率が悪いという問題がある。

(2)においては、可能な文型の全てのアキメントを比較するために比較回数が極めて多いという問題がある。

加えて、両者に重大な問題は、アキメントの用言との相対位置を持つ言語学的意味を見落としていることである。用言との相対位置は、陳述の主体、シフトの可能性、アキメントの新旧性の呈示や、外延の限定という役割を持っている。これらの役割の制御は、相対位置を可能な文型全てに渡って処理できて初めて可能になることである。

従って効率面だけではなく言語学的性質をも容易に反映させることができる新しい行々の開発が必要とされていた。

3. 文型処理行々における課題

文型処理行々は、文の表層と概念依存構造とを対応付けるために、用言と概念素との対応、表層部分と文法格との対応、及び文法格と深層格との対応を正しく関係付けるものでなければならない。尚、文法格を介在させる必要があるのは代名詞の格変化・シフトといった形態・統語的振舞いに加えて意味的制限といった役割を担うからである。

4. 文法格行々

英語における文法格行々で表した文型(文法格行々)を全て列挙すると表1に示すように18行々になり[2,3]、この表から下の二つの性質が言える。

表1. 英語における文法格行々

Table with 2 columns of '用言位置' and 4 columns of grammatical categories (S, IOB, DOB, PP, ADVP, COMP, CONJ). Rows 1-9 and 10-18.

S:主語 IOB:間接目的語 DOB:直接目的語 COMP:補語 ADVP:副詞的小詞 PP:前置詞句 CONJ:接続詞節

Table with 8 columns: S, 用言, IOB, DOB, ADVP, PP, COMP, CONJ.

図1. 英語における文法格の語順

(性質1) 文型を構成する文法格アキメントの個数は最大4である。

(性質2) 文法格アキメント間に図1に示す一貫した順序性がある。

加えて、一般的に、受身シフト、Heavy NP シフト、副詞的小詞シフトが用言直後の位置からしか可能でないこと、用言の前に位置するアキメントは旧情報を表し、用言を含めて後ろは新情報を表すこと、等の言語事実により次の性質が言える。

(性質3) シフト、新旧性、外延、等の性質と文法格アキメントの位置との間には密接な関係がある。

以上3つの性質から文法格の用言との相対位置及び順序に着目して複数の文型を並列処理して絞り込み、表層と構造と意味とを対応付ける行々を開発した。

5. 文型処理モデル

全ての文型は4つのボジション(P1,P2,P3,P4)における文法格アキメントの有無及び種類により表現でき、又、複数の文型に渡って、同じ意味を持つボジションのみを見ることにより並列処理が可能である。

ボジションで表現された文型において文型をルーム、各文法格を窓に見立て、各窓に格マカ、構造型、意味制約、シフト等の言語表層素性を記述することにより、文の表層部分と文法格アキメントとの対応付けが可能となる。

標準語順の表層とは左から順にボジションとの対応が付き、又、シフトの結果としてのアキメントは、キヤップを残すものはトレス規則によりキヤップを起こし得るボジションとの対応付けが可能であり、残さないものは文型を増やすことで対応付けが可能である。

以上のことから複数の文型を処理する場合においても、

↓用言位置

	P1	P2	P3	P4
S	の素性			
S	の素性	COMPの素性		
S	の素性	PPの素性		
S	の素性	DOBの素性		
S	の素性	IOBの素性	DOBの素性	
.....
S	の素性	IOBの素性	DOBの素性	COMPの素性

図2. 英語スロットテーブル

各文型をボジション及びスロット表現し並べること(スロットテーブル)により文法格アキメントと表層部分との対応がつく。

図2の2行目、P2が「COMPの素性」となっているのは、2番目に可能な文型が用言直後にCOMPとしてのアキメントを有し、その素性が知所に記述されていることを示している。空白は文型がそのボジションにアキメントを所有せず、スロットが閉じられていることを示している。

主語ボジション(P1)を除いて各ボジションには複数の文法格が対応しているため、ボジション情報のみではアキメントの文法格を対応付けられない。従って、文型の各々に文法格の番号を付与し、上記(性質2)の文法格の順序性を利用して記述した文法格パターンを用いて、対応する文法格パターン上のボジションにより文法格を対応付ける。

図3において番号10のDOBが1となっているのは、10と番号が付与された文法格がDOBを標準語順において2番目に取ることを示している。

対応する文法格は各知所に表層素性と共に記述し、又、文法格に対応する深層格も辞書情報により対応付けて知所に記述する。

用言表層と意味との関係はどちらから見ても1対多の関係であるが、冒頭で述べたように、両者は文型と密接な関係があり、可能な文型を絞り込むことにより、同時に両者の関係をも絞り込むことができる。

解析においては、スロットテーブルの形式で表現された文型の各々に意味を対応させ、ボジションの一つを選んで文法格の表層素性が文の表層部分の素性と合致する文型のみを妥当なものとし、ボジションを順にずらして同様の処理をすることにより文型を求めると同時に意味を対応付けることができる。文型を求めるときには知所が最も多く合致したものを選ぶ。

生成においては、用言意味の代わりに表層を文型の各々に対応させ、文法格の表層素性の代わりに深層素性(深層

番号	S	IOB	DOB	ADVP	PP	COMP	CONJ
1	1	0	0	0	0	0	0
2	1	0	0	0	0	1	0
6	1	0	0	0	1	0	0
9	1	0	0	0	0	0	1
10	1	0	1	0	0	0	0
11	1	1	1	0	0	0	0
12	1	0	1	0	1	0	0
16	1	0	1	0	0	1	0
17	1	1	1	0	0	1	0

(1: 取る, 0: 取らない)

図3. 文法格パターンテーブル

格、主題性、新旧性、外延、等)を条件として概念依存構造のデバインメントの素性と比較する。解析と同様に文型を求めると同時に用言表層を対応付ける。

文型が求まることにより文法格パターンが対応し各アキメントに対応する文法格が求まる。

6. 既存のワイルドとの比較

既存のワイルドと較べ次のような統語的・意味的性質を反映させることができる。

ソト: 前述のように受身ソト、Heavy NP ソト、副詞的小詞ソトを受けるアキメントは全て P2 に対応付けることができる。

新旧性: 旧情報は P1 に設定することができる。例えば、「ジョンは騙し易い」におけるジョンは旧情報を表し、下の a の構文を選択できる。

a. John is easy to deceive.

b. It is easy to deceive John.

主題性: 主題性を持つアキメントも同様に P1 に設定することができる。例えば、「象は鼻が長い」における「象」は主題を表し、下の c を選択できる。

c. An elephant has a long nose.

d. An elephant's nose is long.

外延: 外延の狭いアキメントを P1 や P2 に設定できる。P2 は受身主語と IOB の場合である。例えば、「ある2カ国語が誰も喋れる」における「2カ国語」は定まったものであり、下の e を選択できる。

e. Two languages are spoken by everybody.

f. Everybody speaks two languages.

又、外延の狭い着点格を P2 における IOB に選ぶことができる。下の例において外延の狭い her を P2 に選び g を選択できる。

g. I gave her a book.

h. I gave a book to her.

全体性: 全体(holistic)の素性を持つ深層格アキメントを P2 における DOB に選ぶことができる。下の例において壁(the wall)全体が塗られたならば i を選択できる。

i. He smeared the wall with paint.

j. He smeared paint on the wall.

7. 終わりに

文表層と概念依存構造とを結ぶ鍵となる文型処理法について述べた。当ワイルドは、複数の文型に渡り一つの共通した意味を持つボジションのみを比較する効率的な並列処理方式であると同時に、言語の統語・意味的振舞いを反映した処理をも可能としている。

既に当ワイルドは、商品化されている英語解析・生成システム、及びスイン語生成プロトタイプシステムに採用され、その動作が確認され、その有用性が確認されている。従って、構造語や孤立語のような基本語順と規則的なアキメント移動を特徴とする言語に対して適したワイルドである。

参考文献

- [1] 田中: 「解析から合成までを融合した英日機械翻訳システム」、日経エレクトロニクス、1983.8.29、pp.275-293。
 [2] A. S. Hornby, Guide to Patterns and Usage in English, 2nd ed., Oxford University Press
 [3] A. P. Cowie, R. Mackin, & I. R. McCaig, Oxford Dictionary of Current Idiomatic English, Oxford University Press
 [4] 末松、福持: 「PIVOT J-E: 英語生成の概要」、昭和63年電子情報通信学会秋期全国大会予稿集
 [5] 野村: 「機械翻訳システム PIVOT における格パターンの処理」、情報処理学会第38回全国大会、1989年3月